Amendments to the Specification

Please replace the title beginning at page 1, line 5, with the following rewritten title:

~~SYSTEM AND METHOD FOR~~ PROVIDING CAPITALIZATION CORRECTION FOR UNSTRUCTURED EXCERPTS

Please replace the paragraph starting at page 1, line 8 with the following rewritten paragraph:

The invention relates in general to text capitalization correction and, in particular, to ~~a system and method for~~ providing capitalization correction for unstructured excerpts.

Please replace the paragraph starting at page 1, line 20 with the following rewritten paragraph:

Web content is relatively unstructured in terms of grammar and standardized usage. Web content is often presented in the form of excerpts, which are primarily short, self-contained narratives including one or more headlines and accompanying text. Excerpts might occur as an artifact of the graphical nature of the Web, which emphasizes the tabular presentation of information. In addition, grammatical rules are often ignored in Web content, which can be typified by incomplete sentences, improper capitalization and often bad prose.

Please replace the paragraph starting at page 2, line 8 with the following rewritten paragraph:

Third party advertisers, in particular, can be at odds with editorial guidelines, yet can benefit by advertising on-line. Compliance is important because the Web provides a vehicle to ~~inexpensively~~ reach a potentially large audience inexpensively. Advertisements can be provided with existing Web content, such as in conjunction with on-line news and information. Advertisements can also be tied to results generated by search engines to build on the topical nature of the underlining query.

Please replace the paragraph starting at page 2, line 21 with the following rewritten paragraph:

Conventional approaches to ensuring compliance with editorial guidelines and similar requirements often employ manual or rote correction of word capitalization. However, such approaches can be ~~slower~~ slow, time-consuming and expensive. Moreover, blanket capitalization correction can overcompensate by removing non-standard and "unusual" forms of acceptable capitalization, such as found in certain proper nouns. For instance, "PlayStation" is a properly capitalized registered trademark. Blanket capitalization correction can be particularly impractical for a large number of product or service advertisements.

Please replace the paragraph starting at page 2, line 29 with the following rewritten paragraph:

Therefore, there is a need ~~for an approach~~ to ~~providing~~ improve capitalization correction of words identified in excerpts from, for instance, Web content. Preferably, such an approach would enforce grammatical and editorial guideline conventions and would accommodate frequently occurring yet non-standard capitalization variations.

Please replace the paragraph starting at page 3, line 3 with the following rewritten paragraph:

There is a further need for ~~an approach to bringing~~ generating a lexicon containing capitalization variations for use in capitalization correction. Preferably, such ~~an approach would~~ generation should facilitate grammatical and editorial guideline compliance.

Please replace the paragraph starting at page 13, line 12 with the following rewritten paragraph:

FIGURE 7 is a flow diagram showing the routine 90 for aggregating a lexicon 42 for use in the method 70 of FIGURE ~~6~~ 5. One purpose of this routine is to produce the lexicon 42, containing non-standard capitalization variations 53 of individual words 52 identified in unstructured excerpts 66. The capitalization variations 53 are selected based on statistical significance and preferably excluded from the implicit rules 41. The routine 90 is described as a sequence of process operations or steps, which can be executed, for

instance, by the aggregator 34 of FIGURE 2, or other components.

Please replace the paragraph starting at page 15, line 5 with the following rewritten paragraph:

~~Each~~ An excerpt 66 is selected (block 111) and the individual words and punctuation marks within the excerpt 66 are tokenized (block 112). A word includes any sequence of characters appearing in a contiguous order or connected by an express grammatical connector, such as a hyphen or underscore. One or more of the tokenized words is iteratively processed (blocks 113-122), as follows. The tokenized word is first examined (block 114). If the tokenized word contains a number or lacks vowels (block 115), the tokenized word is skipped. In English, words containing a number or which lack vowels are generally non-standard words such as found, for instance, in product serial numbers. However, in other languages, word composition could vary and tokenized words containing a number or lacking vowels could be allowed. Otherwise, the tokenized word is optionally matched to the non-standard capitalizations maintained in the lexicon 42 (block 116). If the tokenized word is found in the lexicon 42 (block 117), the best matching word form found in the lexicon 42 is used (block 118). Otherwise, if the tokenized word is "small" and does not occur at the start of the phrase (block 119), the tokenized word is provided in lowercase (block 120). In English, a "small" word includes those words that are ordinarily not capitalized when appearing in a title, even though other words may be capitalized. Small words include articles ("a, an, the"), conjunctions ("and, but, or, nor"), and prepositions shorter than five characters ("as, at, by, for, etc."). However, in

other languages, small words can include other types of words and articles, conjunctions and prepositions shorter than five characters could be allowed. In addition, other non-small words could be provided in lower case. Otherwise, the first letter of the tokenized word is capitalized and the remaining letters provided in lowercase (block 121). Processing continues with each remaining tokenized word in the excerpt 66 (block 122), after which the routine returns.

Please replace the Abstract beginning at page 25, line 4, with the following rewritten Abstract:

~~A system and method for providing~~ Providing capitalization correction for unstructured excerpts is described. An excerpt of unstructured content is tokenized into a set of words. The set of words is analyzed for correct capitalization. Individual characters constituting at least one such word in the set of words are evaluated. The at least one such word is skipped if determined to be of a predefined type.